Deer Hunting Mathematical Model

Taylor Larrechea

Colorado Mesa University

Department of Mathematics and Statistics

May 3, 2018

"Brilliant. Got another bag?"

Abstract

Imagine a Parks and Wildlife department is trying to make estimates for how many deer tags to issue to registered hunters for one giving year. This same department aims to predict how many deer will be killed knowing how many hunters there are as well as the allotted total recreation days among all units in the state. The purpose of this mathematical model is to provide an estimate for the deer population killed after a hunting season so the Parks and Wildlife departments can know how many tags to issue each year. This model takes all matters of take into account, i.e the way the deer was killed and compares these totals to other variables that fluctuate from season to season. Two separate models were derived, both depending upon one variable each, and then one final model was constructed.

1 Problem Statement and Introduction

The purpose of this mathematical model is to predict the harvest from deer hunting in any given year. The data collection for this model came from Colorado Parks and Wildlife [1] and spanned from 2004-2013. Since this model persists mainly of data collection and fitting curves to the data, this model falls into the Empirical Model category. There are several assumptions that are made in the construction of this model before we can begin constructing it. Two separate models were made to predict the harvest from deer hunting in any given year and the two are later compared with one another. A final model is constructed with both variables in the equation.

2 Assumptions and Variables

This mathematical model aims to look at the impact of specific variables on deer population during hunting season. For starters P_i and P_f both represent the initial and final populations of deer for one season respectively. The primary variable that we are looking at is T_H an it represents the total harvest of dear in a certain year. This total harvest value depends on the total number of hunters T_h for one model. Where as the total harvest depends on the total number of recreation days T_R in a separate model before the two variables are combined into one model. The primary assumption that we are making in this mathematical model is that no deer are being killed off by illegal hunting, or also known as poaching. Since the data that we used to construct this Empirical model only has data for deer harvest from legal manners of take, this is why we make this assumption. We are also assuming that the total number of hunters do not change as well as the total number of allowed recreation days do not change as the hunting season goes on. With these assumptions and definitions we can begin to construct our models.

3 Model Construction

The first mathematical model that is constructed is a model that only takes into account the total number of hunters to predict the total harvest. Data was collected from the Colorado Parks and Wildlife [1] and can be seen in Table 1 below.

Year	P_i (Deer)	T_h (Hunters)	T_H (Deer)	P_f (Deer)
2004	642643	91646	41743	600900
2005	655115	91757	41665	613450
2006	657994	97286	45234	612760
2007	583796	98283	45026	538770
2008	502312	86245	35552	466760
2009	494442	78536	33922	460520
2010	465164	78603	34768	430396
2011	451167	76445	33217	417950
2012	441096	73705	33086	408010
2013	423601	74233	32941	390660

Table 1: General Data

The data from Table 1 was then used to create plots to fit different degrees of polynomials. T_H was graphed against T_h and the linear plot can be seen in Figure 1.

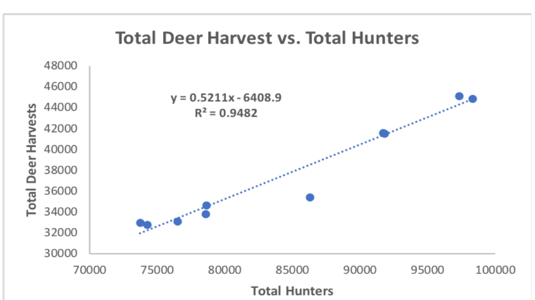

Figure 1: Linear Polynomial

From Figure 1 we can see an equation that was fit to the data. This equation will serve as the linear model when only taking in the total number of hunters into account to predict total harvest. Equation (1) is

$$T_H = 0.52 \cdot T_h - 6409 \tag{1}$$

now giving us a linear model for predicting the total harvest. The same data was then used to create a plot to fit a second degree polynomial to the data. Figure 2 shows this plot below.

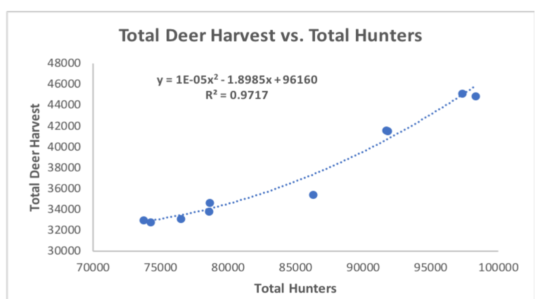

Figure 2: 2^{nd} Degree Polynomial

From Figure 2 we now have our second degree polynomial model. Equation (2) is thus

$$T_H = 1.4 \cdot 10^{-5} \cdot T_h^2 - 1.90 \cdot T_h + 96160 \tag{2}$$

where we now have our second degree polynomial model for predicting total harvest of deer. The same plot was then rendered to fit a third degree polynomial to the data. Figure 3 shows this plot.

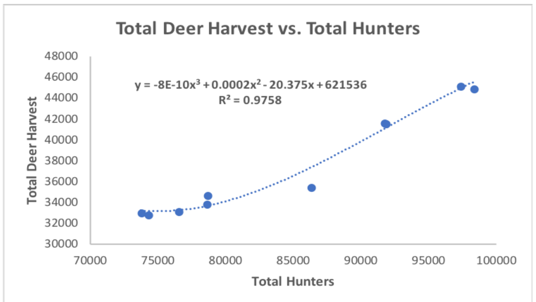

Figure 3: 3^{rd} Degree Polynomial

Figure 3 has another equation with it than can be used to predict the total harvest of deer. This equation is

$$T_H = -8.3 \cdot 10^{-10} \cdot T_h^3 + 2.3 \cdot 10^{-4} \cdot T_h^2 - 20.4 \cdot T_h + 621536 \tag{3}$$

where equation (3) represents the third degree polynomial model. The last polynomial to look at for this model can be seen in Figure 4.

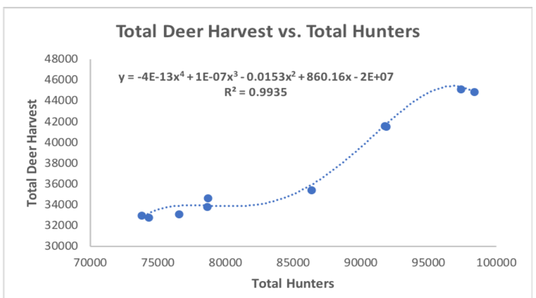

Figure 4: 4^{th} Degree Polynomial

Figure 4 has the final equation that will be used to predict total harvest. The fourth degree polynomial that represents the fourth degree model is

$$T_H = -3.6 \cdot 10^{-13} \cdot T_h^4 + 1.2 \cdot 10^{-7} \cdot T_h^3 - 0.015 \cdot T_h^2 + 860 \cdot T_h - 1.8e^7. \tag{4}$$

Figure 4 and equation (4) are both the last plots and equations that were fitted to the data in Table 1. It should be reiterated that Figures 1-4 and Equations (1)-(4) all only depended upon the number of hunters to help create equations to fit the data. Now we need to do the same thing but with a different variable to predict the total deer harvest. This variable is the T_R variable that represents the total recreation days in a hunting season.

Throughout Colorado there are specified units in the state that all contribute to a total called the, "Recreation Days" which are essentially the days that hunters spent hunting in one season. Table 2 has the data that was used to create the preceding four models where the total recreation days is substituted for the total hunters in Table 1.

Year	P_i (Deer)	T_R (Rec. Days)	T_H (Deer)	P_f (Deer)
2004	642643	270983	41743	600900
2005	655115	394339	41665	613450
2006	657994	440296	45234	612760
2007	583796	435070	45026	538770
2008	502312	395479	35552	466760
2009	494442	364189	33922	460520
2010	465164	360048	34768	430396
2011	451167	338983	33217	417950
2012	441096	332573	33086	408010
2013	423601	338132	32941	390660

Table 2: General Data

Using the data from Table 2, we graphed the total harvest against the total recreation days and fit polynomials to each plot. Immediate impressions were that these models (The total recreation days dependent models) are less reliable and not as accurate as the previous four. Figure 5 shows the linear model for the first total recreation day dependent model.

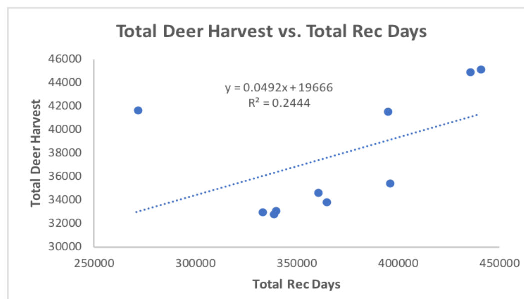

Figure 5: Linear Polynomial

Figure 5 has the equation for the linear recreation days dependent model. The equation for this model is

$$T_H = 0.049 \cdot T_R + 19667 \tag{5}$$

where equation (5) when comparing with equation (1), which was the linear total hunters dependent model, the recreation days dependent model looks to be less reliable right away. We now move to modeling the second degree polynomial for the total recreation days dependent model. Figure 6 shows the second degree polynomial fit to data from Table 2.

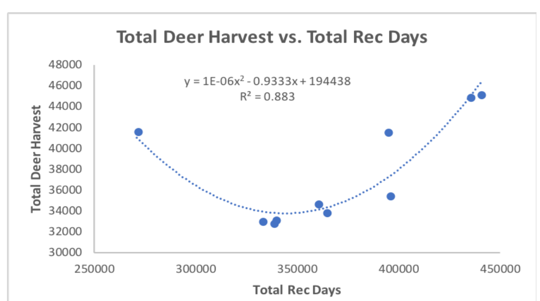

Figure 6: 2^{nd} Degree Polynomial

We then can do the same thing with different degree polynomials again. The equation that describes the second degree polynomial recreation days dependent model is

$$T_H = 1.4 \cdot 10^{-}6 \cdot T_R^2 - 0.93 \cdot T_R + 194438 \qquad (6)$$

where equation (6) can be immediately seen as more accurate (By looking at the R^2 value on the plot) than equation (5). Comparing equation (6) with equation (2), the total hunter dependent models still look to be more reliable for predicting total harvest in comparison with the data set. We now take a look at the third degree polynomial recreation dependent model. Figure 7 shows the data from Table 2 fit to a third degree polynomial.

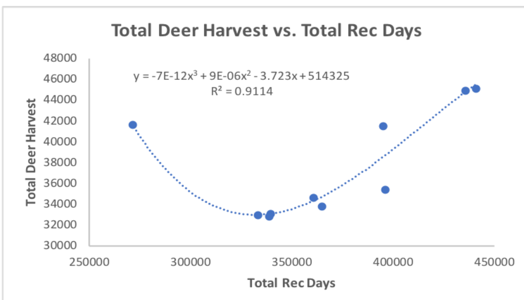

Figure 7: 3^{rd} Degree Polynomial

The same process is done with Figure 7 as with Figure 5 and 6 to find the third degree polynomial model equation. Equation (7) is

$$\begin{aligned} T_H = & -7.5 \cdot 10^{-12} \cdot T_R^3 + 9.38 \cdot 10^{-6} T_R^2 \\ & - 3.7 \cdot T_R + 514325 \end{aligned} \qquad (7)$$

giving us the third degree recreation days dependent model. This model is still not as accurate as the total hunters dependent models. Because of this, we try one more time with a fourth degree polynomial fitting to the data in Table 2.

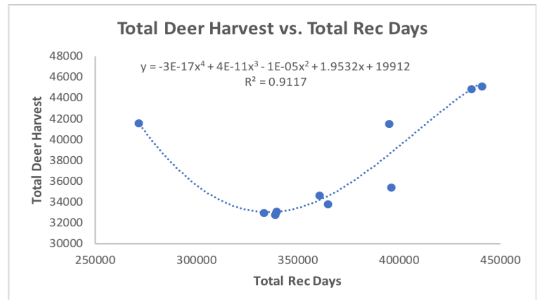

Figure 8: 4^{th} Degree Polynomial

We then have one last equation from Figure 8. Equation (8) is

$$\begin{aligned} T_H = & -3.0 \cdot 10^{-17} \cdot T_R^4 + 3.8 \cdot 10^{-11} \cdot T_R^3 \\ & - 1.5 \cdot 10^{-5} \cdot T_R^2 + 2.0 \cdot T_R + 19912, \end{aligned} \qquad (8)$$

where we now have the last equation for the fourth degree polynomial recreation days dependent model. Now that all of the equations have been fit to the data, we can focus on looking at which model is the most accurate.

4 Model Selection

To recap equations (1)-(4) and Figures 1-4 were total hunters dependent models. Equations (5)-(8) and Figures 5-8 pertained to the total recreation days dependent models. We now aim to compare the models with the actual data that was collected. For equation (1), Figure 9 shows the comparison between the linear model and the actual data.

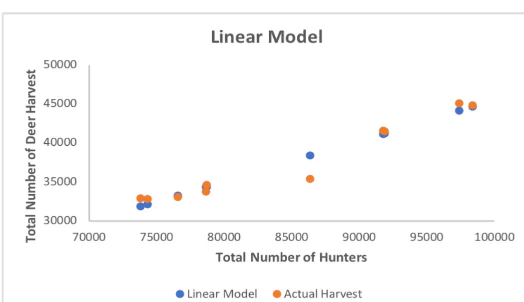

Figure 9: Linear Model Comparison

Equation (1)'s R^2 value was 0.95, so it does a decent job of modeling the data. Table 3 shows the results of the linear model.

Model (Deer)	PRE (%)	Residuals (Deer)
41348	0.95	395
41406	0.62	259
44287	2.09	947
44806	0.49	220
38533	-8.39	-2981
34516	-1.75	-594
34551	0.62	217
33427	-0.63	-210
31999	3.29	1087
32274	2.03	667

Table 3: Linear Model Statistics

Table 3 has PRE's that are relatively small, this indicates that the model does a good job predicting total harvest. Because the errors still fluctuate up to -8%, we still wish to improve this model. We now wish to take a look at the 2^{nd} degree polynomial total hunters dependent model. Figure 10 shows the comparison between the model and the actual data.

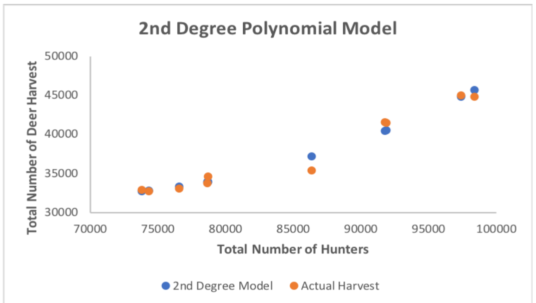

Figure 10: 2^{nd} Degree Model Comparison

Equation (2)'s R^2 value was found to be 0.97, and from the looks of Figure 10 we can already see that the second degree polynomial already does a lot better job than the linear model. We now wish to take a look at the statistics from this model and they can be seen in Table 4 below.

Model (Deer)	PRE (%)	Residuals (Deer)
40671	2.57	1072
40748	2.20	917
44998	0.52	236
45856	-1.84	-830
37369	-5.11	-1817
34082	-0.47	-160
34103	1.91	665
33480	-0.79	-263
32877	0.63	209
32977	-0.11	-36

Table 4: 2^{nd} Degree Model Statistics

Table 4 shows that the second degree polynomial is a lot more accurate than the linear model. We can also observe the distance between the orange and blue dots getting smaller in Figure 10. This is another indicator that the second degree model is better. We continue on and now look at the third degree polynomial total hunters dependent model comparison to the actual data. Figure 11 shows the comparison between the model and the actual data.

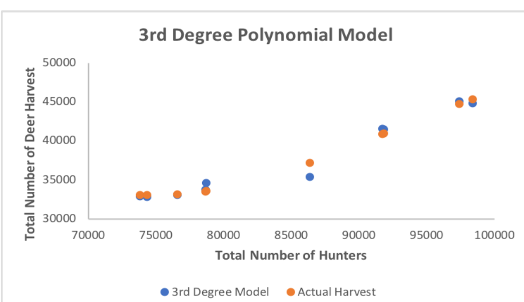

Figure 11: 3^{rd} Degree Model Comparison

Equation (3) was found to have an R^2 value of 0.98. We can also see in Figure 11 that our model is improving by adding in more terms to the polynomial. Table 5 has the statistics for this data.

Model (Deer)	PRE (%)	Residuals (Deer)
41039	1.69	704
41118	1.31	547
44920	0.69	314
45542	-1.15	-516
37343	-5.04	-1791
33681	0.71	241
33699	3.07	1069
33275	-0.17	-58
33259	-0.52	-173
33212	-0.82	-271

Table 5: 3^{rd} Degree Model Statistics

We can again see that the results are improving in Table 5 above. The last model to look at for the total hunters dependent models is the fourth degree polynomial model. Figure 12 shows the comparison of the model with the actual data.

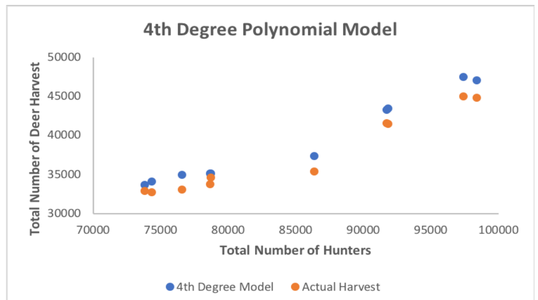

Figure 12: 4^{th} Degree Model Comparison

Equation (4) was used to graph the model in Figure 12 and the R^2 value for this model ended up being 0.99. This model has the highest R^2 value out of all the models. We now wish to look at the statistics for this model in Table 6.

Model (Deer)	PRE (%)	Residuals (Deer)
43505	-4.22	-1762
43640	-4.74	-1975
47726	-5.51	-2492
47305	-5.06	-2279
37563	-5.66	-2011
35285	-4.02	-1363
35285	-1.49	-517
35163	-5.86	-1946
33855	-2.33	-769
34274	-4.05	-1333

Table 6: 4^{th} Degree Model Statistics

It should be noted that even though the fourth degree model has the highest R^2 value, it doesn't mean that it is the most reliable. As can be seen in Figure 4, this model has a lot of fluctuations in it throughout the data indicating that it may not be reliable for data sets outside of what we have. Because of this we won't be using the fourth degree model as the representative total hunters dependent model. Instead, because the third degree polynomial's R^2 value is still relatively high (0.98) we will use this model. Now that the best total hunters dependent model is determined, we can focus our attention now on solving for the total recreation days dependent model.

The linear total recreation days dependent model is a lot more inaccurate than the total hunters dependent model. Looking at equations (5)-(8) we now will look at the accuracy of these models. Figure 13 shows the comparison of the linear total recreations day dependent model to the actual data.

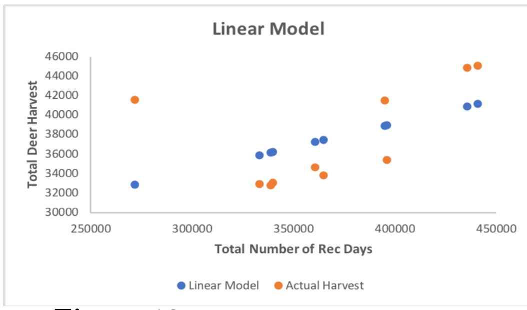

Figure 13: Linear Model Comparison

Equation (5) was found to have an R^2 value of 0.24. We can immediately see from the R^2 value that this model is not doing a good job of predicting total harvest. Table 7 shows the statistics of this model.

Model (Deer)	PRE (%)	Residuals (Deer)
32993	20.96	8750
39060	6.25	2605
41320	8.65	3914
41063	8.80	3963
39116	-10.02	-3564
37577	-10.77	-3655
37373	-7.49	-2605
36337	-9.39	-3210
36022	-8.87	-2936
36296	-10.18	-3355

Table 7: Linear Model Statistics

From Table 7 we can observe that the PRE as well as the residuals for this model aren't very accurate. Because of this we will try again with a second degree model. Figure 14 shows the second degree total recreation days dependent model in comparison with the actual data.

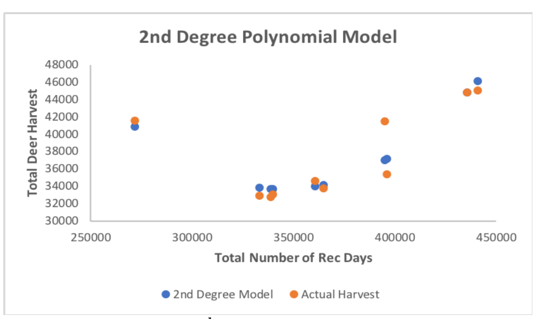

Figure 14: 2^{nd} Degree Model Comparison

Equation (6) was found to have an R^2 value of 0.88. Immediately we can see that there is an improvement from the linear total recreation days dependent model. This model still isn't as accurate as its total hunters dependent counterpart, and the statistics of this model can be seen in Table 8.

Model (Deer)	PRE (%)	Residuals (Deer)
41076	1.60	667
37212	10.69	4453
46322	-2.41	-1088
44998	0.06	28
37369	-5.11	-1817
34347	-1.25	-425
34146	1.79	622
33843	-1.88	-626
33989	-2.73	-903
33856	-2.78	-915

Table 8: 2^{nd} Degree Model Statistics

In comparison to the linear model, this second degree model is a lot more accurate. The accuracy and reliability of the model will improve in the third degree polynomial model. Figure 15 shows the comparison of the third degree total recreation days dependent model and the actual data.

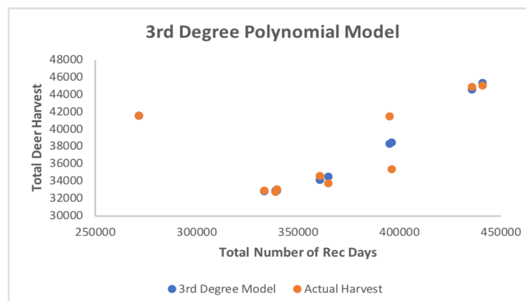

Figure 15: 3^{rd} Degree Model Comparison

Equation (7) was found to have and R^2 value of 0.91. So this model is increasing when the power of the polynomial increases, which is what we can expect. This model still isn't perfect or better than its counterpart. Table 9 has the statistics of this third degree model.

Model (Deer)	PRE (%)	Residuals (Deer)
41755	-0.03	-12
38454	7.71	3211
45487	-0.56	-253
44722	0.67	304
38622	-8.64	-3070
34678	-2.23	-756
34296	1.36	472
33087	0.39	130
33003	0.25	83
33067	-0.38	-126

Table 9: 3^{rd} Degree Model Statistics

The results from Table 9 confirm that the data is improving. The last model to look at in total (Between the total hunters and total recreation days) is the fourth degree total recreation days dependent model. Figure 16 shows the comparison between the fourth degree model and the actual data.

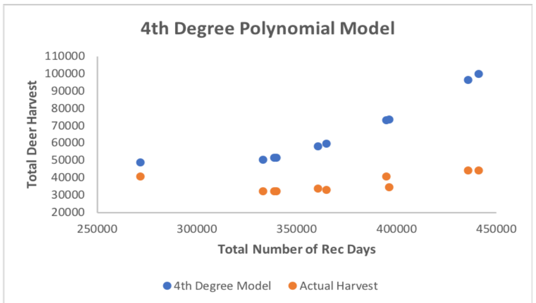

Figure 16: 4^{th} Degree Model Comparison

Model (Deer)	PRE (%)	Residuals (Deer)
49681	-19.02	-7938
74101	-77.85	-32436
100726	-122.68	-55492
97514	-116.57	-52488
74692	-110.09	-39140
60451	-78.21	-26529
58915	-69.45	-24147
52543	-58.18	-19326
51104	-54.46	-18018
52338	-58.88	-19397

Table 10: 4^{th} Degree Model Statistics

The R^2 value for equation (8) is 0.91. This model does not offer much improvement over the third degree model. Table 10 shows the statistics for the fourth degree total recreation days dependent model.

With Table 10 we have our last statistics table to look at before choosing a final model. Out of all of the equations (1)-(8) the most accurate and reliable model was found to be that of equation (3). This is the third degree total hunters dependent polynomial model. The reason why this model wins out is because of its R^2 value and the fact that the trend line doesn't fluctuate as rapidly as others. This model has a PRE at the highest of 5.11% and a residual at the highest of 1817. This model is the clear cut most accurate model out of all eight to select from.

5 Implementing the Model

From the model selection category we deduced that the most accurate model was that of equation (3). The reason why this model was chosen was because the trend line from this degree polynomial made the most sense out of any of the models and had the least amount of fluctuation. Now the task is to present this model in an easy way for a Wildlife Department so they can use it to make predictions. Although equation (3) proved to be very accurate for predicting total harvest, a Wildlife Department is more than likely going to want to incorporate a different variable to help make that prediction. We will now create our final mathematical model that is both dependent on total hunters and total recreation days to predict total harvest. The first step was that regression analysis was done with excel on Table 11 below.

Year	T_H (Deer)	T_h (Hunters)	T_R (Rec. Days)
2004	41743	91646	270983
2005	41665	91757	394339
2006	45234	97286	440296
2007	45026	98283	435070
2008	35552	86245	395479
2009	33922	78536	364189
2010	34768	78603	360048
2011	33217	76445	338983
2012	33086	73705	332573
2013	32941	74233	338132

Table 11: General Data

After the regression analysis equation (9) is created to represent this multiple variable model

$$T_H = 0.54 \cdot T_h - 0.005 \cdot T_R - 5820, \qquad (9)$$

where equation (9) is our final mathematical model for predicting total deer harvest. More precise measurements can be seen in the spread sheet that was made to create this model, but equation (9) was found to have an R^2 value of 0.95. Equation (9)'s R^2 value may be less than equation (3)'s, but equation (9) has more variables in it and thus is why we are choosing it as the final mathematical model. Figure 17 shows the final model comparison when it is graphed along with the actual harvest data.

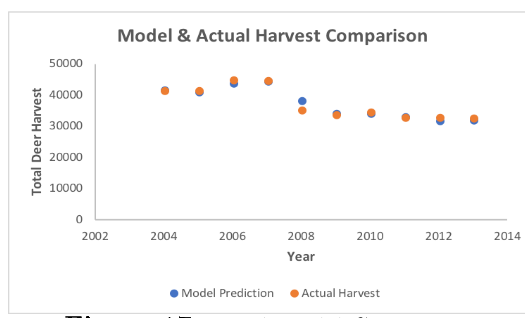

Figure 17: Final Model Comparison

We can see visually that equation (9) does a very good job in comparison to the actual data that was collected from the Colorado Parks and Wildlife. The last step to verify the accuracy of this model is to show the statistics of it. Table 12 compares the model with the actual harvest statistically.

Model (Deer)	PRE (%)	Residuals (Deer)
41922	-0.43	-179
41374	0.70	291
44108	2.49	1126
44668	0.80	358
38417	-8.06	-2685
34442	-1.53	-520
34499	0.77	269
33447	-0.69	-230
32011	3.25	1075
32266	2.05	675

Table 12: Final Model Statistics

We can confirm with Table 12 that this model does a good job at predicting the total deer harvest each year. Figure 18 shows the table that the Colorado Parks and Wildlife will be using to make their predictions. It is important to note that the values highlighted in orange are the total hunters, the values highlighted in green are the total recreation days among all units, and the values highlighted in blue are is the total deer harvest in one year off the proposed values.

	270000	287000	304000	321000	338000	355000	372000	389000	406000	423000	440000
73000	31942	31858	31774	31691	31607	31523	31439	31356	31272	31188	31104
75500	33281	33197	33113	33029	32946	32862	32778	32694	32611	32527	32443
78000	34619	34536	34452	34368	34284	34201	34117	34033	33949	33866	33782
80500	35958	35874	35791	35707	35623	35539	35456	35372	35288	35204	35121
83000	37297	37213	37129	37046	36962	36878	36794	36711	36627	36543	36459
85500	38636	38552	38468	38384	38301	38217	38133	38049	37966	37882	37798
88000	39974	39891	39807	39723	39639	39556	39472	39388	39305	39221	39137
90500	41313	41229	41146	41062	40978	40894	40811	40727	40643	40560	40476
93000	42652	42568	42484	42401	42317	42233	42150	42066	41982	41898	41815
95500	43991	43907	43823	43740	43656	43572	43488	43405	43321	43237	43153
98000	45329	45246	45162	45078	44995	44911	44827	44743	44660	44576	44492

Figure 18: Model Implementation

Figure 18 implements equation (9) and makes it a very user friendly table to use for the parks service who would want to use this model. The concludes the constructing and choosing of a deer harvest mathematical model.

6 Conclusion

We first started off with constructing a mathematical model that depended upon one variable. Two mathematical models were constructed at first, one dependent upon total hunters and one dependent upon total recreation days. The total hunters dependent model came out to be accurate just with one variable, but when the total recreation days was added to it it improved the accuracy of the model. Equation (9) is also a lot simpler than equation (3) making it more user friendly if someone wants to find an exact value for total harvest. Figure 18, which is essentially a table for estimates with no calculations required, is what should be handed to someone wanting to use this mathematical model.

References

[1] Wildlife, Colorado Parks and. Colorado Parks and Wildlife. 1 January 2018. 28 April 2018. ¡http://cpw.state.co.us/thingstodo/Pages/Statistics-Deer.aspx¿.